



Open Preservation

Community approaches to ensuring long-term access to digital collections

The OPF team



Community Manager

- Events (face-to-face/virtual)
- Training (staff development)
- Comms (web/email/social)



Executive Director

- Membership (engagement/value)
- Open preservation advocacy
- Operational management



Technical Lead

- Infrastructure (host/test)
- Software stewardship (roadmap/maturity/packaging)
- Data corpora



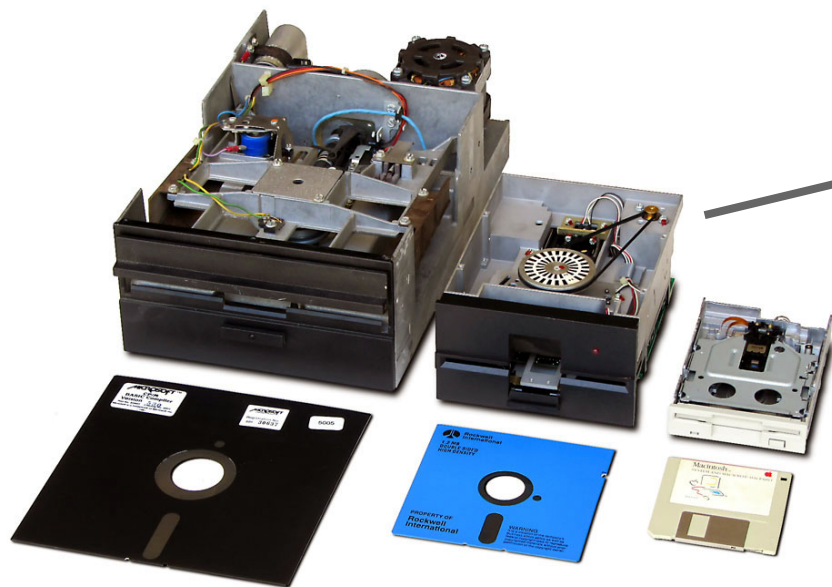
Overview

- Preservation Scenarios: scale and perspective
- State of the Art: software and practices
- Repository Integration: tools and user experience
- The Role of Community: shared knowledge + shared effort == sustainable solutions
- Opportunities for Collaboration

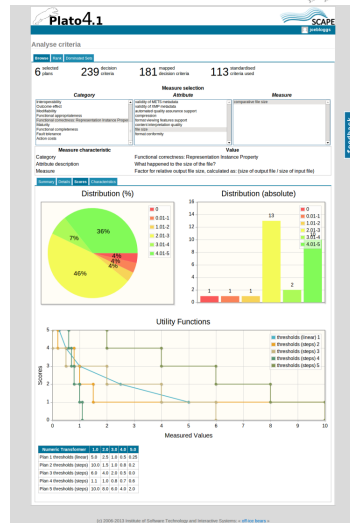
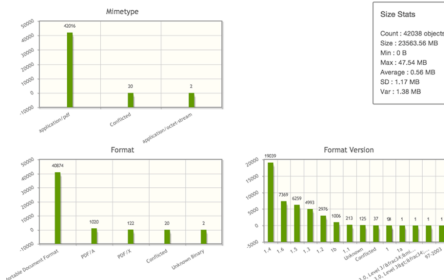
Preservation Scenarios: small scale

- Characterisation and validation
 - Profiling heterogeneous collections (e.g. deposited archives)
 - Analysing disk image contents (e.g. legacy media)
 - Dealing with email formats (e.g. institutional records)
- Detecting duplicates (e.g. collection digitisation projects)
- Automated metadata extraction

Preservation Scenarios: small scale



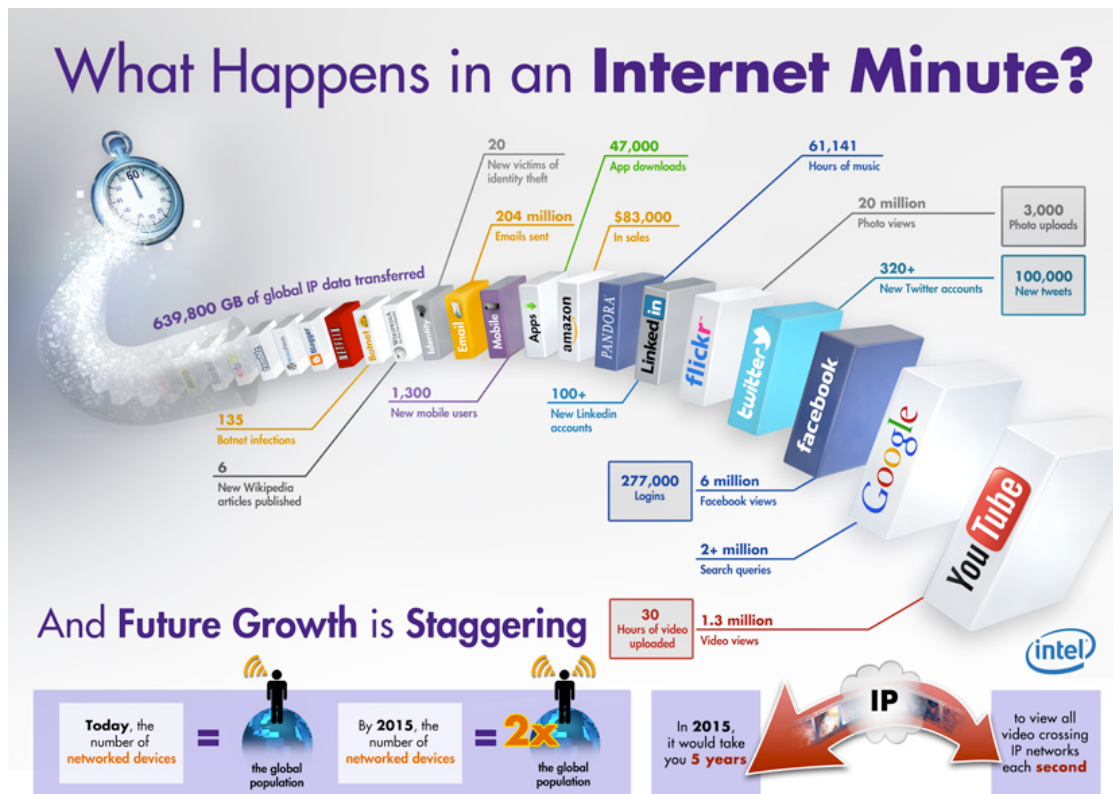
Overview



Preservation Scenarios: med - large

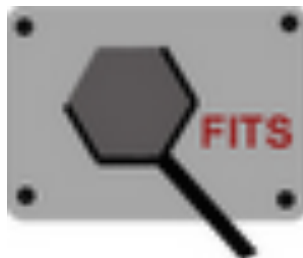
- Automation (e.g. quality control of format migrations)
- Scalability (rapidly growing collections)
- Diverse sources (e.g. validating 3rd party submissions)
- Complexity (variety of formats and files, e.g. digital rights management, digital art, or non-conformance to specs)

Preservation Scenarios: the future?



State of the Art: common tools

- Identify
- Characterise
- Validate
- Analyse



BitCurator

The BitCurator logo features a blue sphere with a white satellite or probe icon orbiting around it.

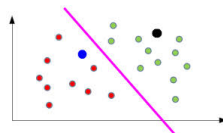
JHOVE2

The JH logo is a grey sphere with horizontal lines, resembling a planet or a globe.

State of the Art: SCAPE tools

Characterisation and Quality Assurance:

- pagelyzer - web page rendering
- matchbox - image analysis (duplicates, cropping)
- jpylyzer - JPEG 2000 validator [jpylyzer](#)
- xcorr sound - audio waveform analysis

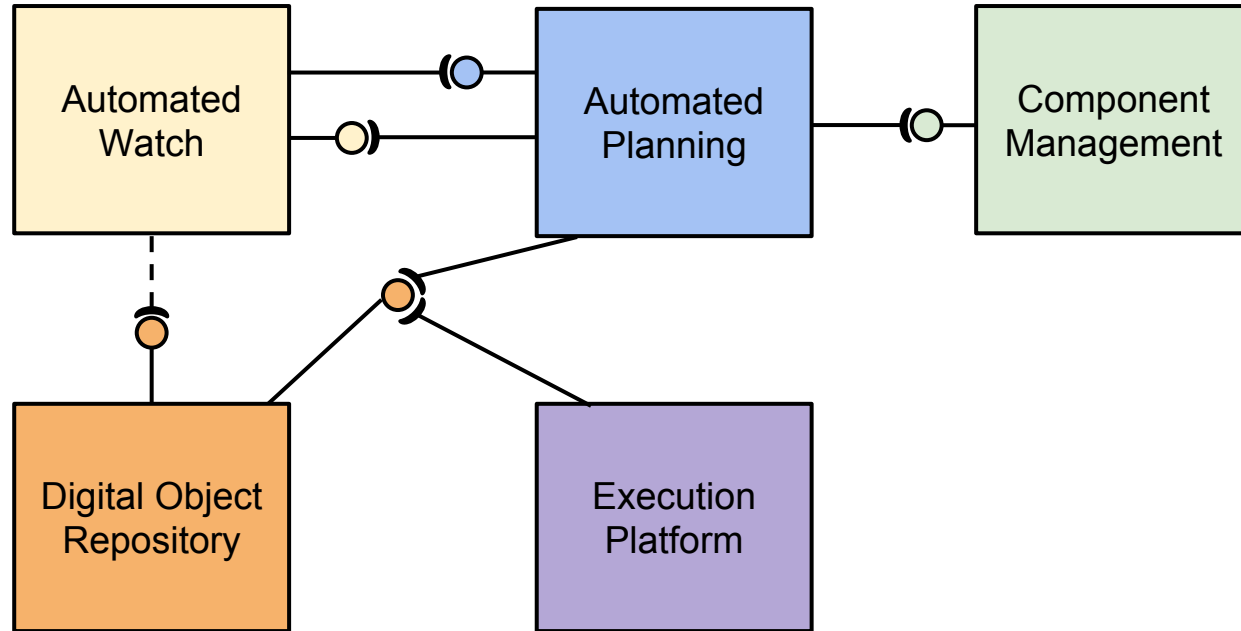


State of the Art: SCAPE systems

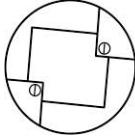



- PLATO (preservation planning)
- SCOUT (preservation watch)
- Platform (parallel processing)
- C3PO (content profiling)



State of the Art: SCAPE ecosystem



Repository Integration

- RODA (Fedora 3)  **RODA**
REPOSITORY OF AUTHENTIC DIGITAL OBJECTS
- Fedora 4 (reference SCAPE integration)  **Fedora™**
- Rosetta (Ex Libris) 
- Hydra (digital preservation working group) 

Repository Integration

- Eprints (preservation plugins)
- Preservica (Tessella)
- microservices / workflows



Repository Integration: make it easy

Improving usability of digital preservation tools

- Discoverability (easy to find)
- Packaging (easy to install)
- Documentation (easy to use)

Community: Open Knowledge

Shared Knowledge

- reports
- case studies
- blogs
- wikis

Connecting digital preservation expertise

Home / Community

Blogs

Developing an Audio QA workflow using Hadoop: Part II

First things first. The Github repository with the Audio QA workflows is here: <https://github.com/statsbiblioteket/scape-audio-qa>. And version 1 is working. Version is really all wrong here. I should call it Workflow 1, which is this one:

Submitted by [Bolette Ammitzb...](#) on 3 February 2014 - 11:35am

Welcome to the OPF Knowledge Base Wiki

The OPF wiki is here to gather knowledge relating to key issues in digital preservation and organise these to allow practitioners, developers and information specialists to better preserve digital heritage. This wiki welcomes contributions from every member of the community and provides a hosting space for many global projects. From practitioners, we welcome your use case scenarios and problems and hope to link these with solution providers. For developers, we not only hope to join you to practitioners in the community through online and face-to-face events, but also welcome your valuable solutions with advice on software development, project hosting and distribution. Gathering together communities and key knowledge allows **information managers** to make effective decisions for preserving our digital heritage.

- Contribute your requirements and solutions**
Share your challenges and solutions with the community, and view the resulting record of digital preservation practice
- Collaborate**
Get involved with these collaborative digital preservation initiatives, including Q&A on Stack, the Atlas of Digital Damages and the OPF Format Corpus
- Find preservation tools**
Find tools for solving your digital preservation challenges in the OPF Tool Registry
- Understand File Format Risks**
File format risks you should be concerned about, and how to identify them in your content
- Develop Software with the OPF**
All you need to know about developing software with the OPF, with guidance on making your software effective, sustainable, maintainable and reusable
- OPF Projects**
Projects hosted by OPF, including: [AQuA](#), [SCAPE](#), [SPRUCE](#) and [Preservation Health Check](#)
- OPF Events and Webinars**
The latest OPF events and hackathons
- Join the conversation**
Blog, comment and exchange ideas on the main [Open Planets Foundation website](#)

Recent comments

- Survey closed**
Thank you all for your participation! The survey was a success with more than 350 participations....
[Luis Faria](#) 1 week 4 days ago
- Accessing the API through Python**
On a related note: the other week I wanted to get a more detailed look at Tika's mimetyp detection...
[Johan van der Knijff](#) 1 week 6 days ago
- Last call and thanks**
Last call for the survey participation! Survey will close at the end of the week, February 28th, so...
[la](#) 4 weeks 18 hours ago
[S](#), amongst others...
:h noting that 'Java first' is also the approach es, and in general FITS seems to...
[ckson](#) 4 weeks 21 hours ago
a program using Java libs
s, Very interesting post. An alternative for speeding up execution would be to...
[Palmer](#) 4 weeks 23 hours ago
:o be most concerned about security?
Thanks for the comment and the nent on the post. Asking from a rather naive '...
[encer](#) 4 weeks 1 day ago
security
Interesting stuff here, I've looked at a few times. Was unaware of the memory...
[son](#) 4 weeks 1 day ago
w/ your idea Andrea...
his goes back to the problem of governance. mentioned tools have different owner,...
[ar](#) 1 month 1 week ago

Sign up for the OPF wiki

Datasets, Issues and Solutions

OPF Format Corpus

Atlas of Digital Damages

Open Planets Foundation

Community: Open Knowledge

Community Owned digital Preservation Tool Registry (COPTR)

COPTR describes tools useful for [preserving digital information for the long term](#). COPTR is also an initiative to collate the knowledge of the digital preservation community on preservation tools in one place. Instead of organisations [competing against each other with their own registries](#), COPTR is bringing them together. In doing so its objective is to provide the best resource for practitioners on digital preservation tools.

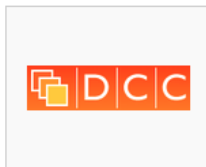
Browse the COPTR Registry

- [View all tools](#)
- [View tools by functional category](#)
- [View tools by the type of content they act upon](#)

How to create a new Tool Entry

1. Check out these [Guidelines for contributing to COPTR](#).
2. [Search for the tool you want](#). Check the full name and the acronym!
3. If you find it, consider adding more detail to the existing entry. If you don't find it, follow the link to create a new entry, which will automatically create an outline entry for your tool based on the [Tool Page Template](#).

COPTR Partners



The Digital Curation
Centre (DCC) [↗](#)



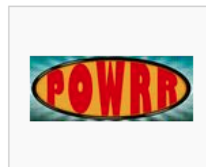
The Digital Curation
Exchange (DCE) [↗](#)



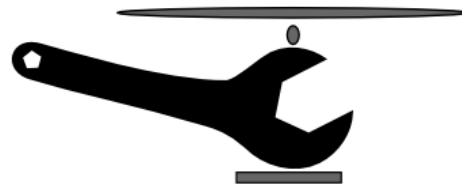
National Digital
Stewardship Alliance
(NDSA) [↗](#)



The Open Planets
Foundation (OPF) [↗](#)



Preserving digital Objects
With Restricted Resources
project (POWRR) [↗](#)

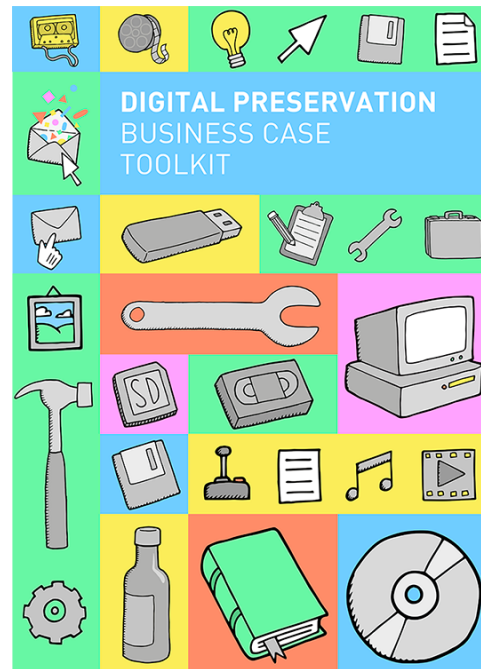


Community: Open Knowledge

Digital Preservation

Business Case Toolkit

- making the case
- benefits, risks, costs...
- templates, FAQ, case studies



Community: Open Knowledge

Hack Events:

- Traditionally face-to-face (developers and practitioners)
- Start with a tell and show / speed dating format
- Followed by development on specific issues



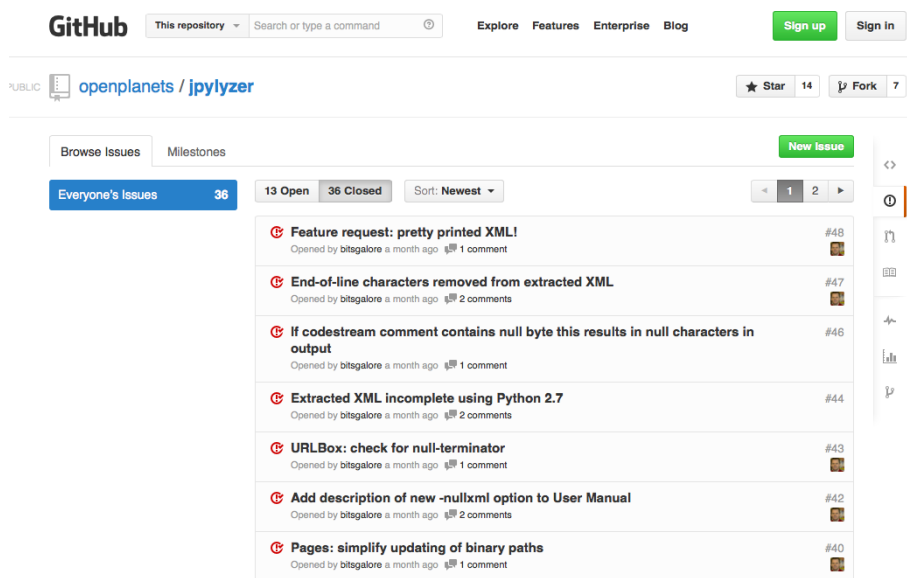
Community: Open Software

- Shared Requirements

- Wiki Pages
- Issue Logs
- **Testability**

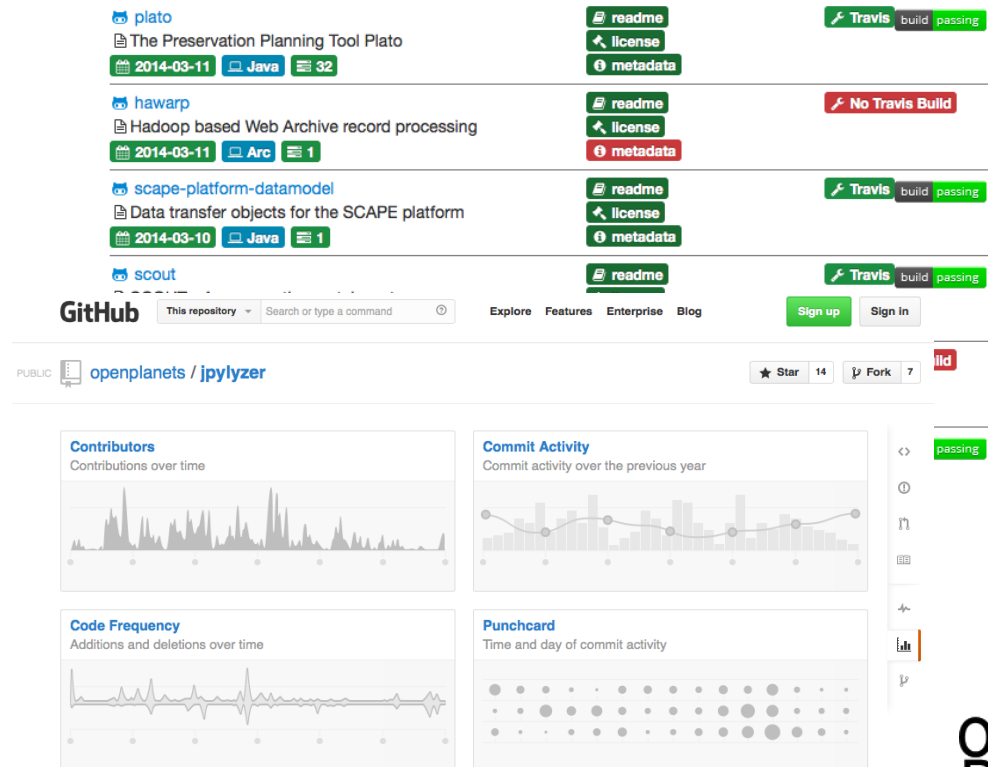
- Shared Experiences

- Wiki pages
- Blogs



Community: Open Software

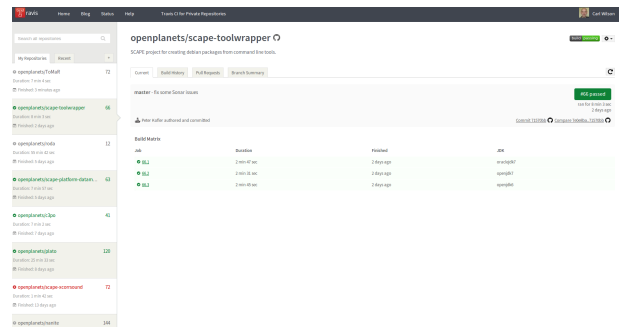
- Shared Code
 - Open Source
 - Visibility
 - Project quality



Community: Open Software

- Shared Testing == Robust Tools

- Continuous integration
- Community corpora
- Functional testing
- Regression testing
- Published results
- Towards continuous deployment



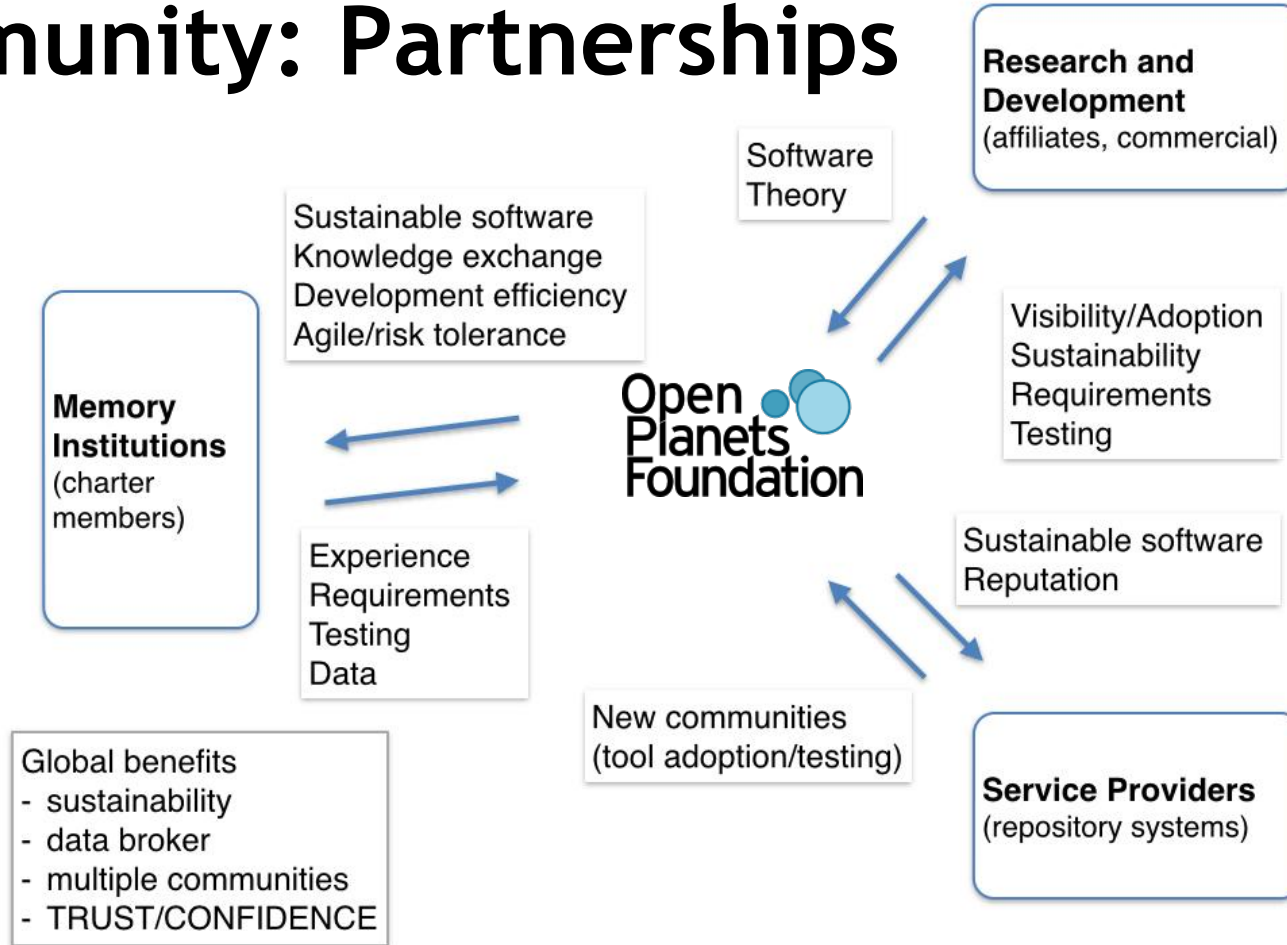
Community: Partnerships



THE ROYAL LIBRARY
NATIONAL LIBRARY OF DENMARK AND COPENHAGEN UNIVERSITY LIBRARY



Community: Partnerships



Opportunities for Collaboration

- Digital preservation can seem technically complex, what are the needs for integrating available tools and services with widely adopted repository systems?
- Solutions can appear achievable only at scale (large institutions with dedicated teams), what are the requirements for deployable solutions at all scales of operation?
- The repository and preservation communities can seem isolated from each other, how can we ensure better collaboration on shared challenges?

My Details

- Carl Wilson
- Technical Lead
- Open Planets Foundation
- Email : carl@openplanetsfoundation.org
- Skype : [carl.f.wilson](https://www.skype.com/people/carl.f.wilson)
- GitHub : [carlwilson](https://github.com/carlwilson)
- Twitter : [@openplanets](https://twitter.com/openplanets)
- Google+ : [carl@openplanetsfoundation.org](https://plus.google.com/carl@openplanetsfoundation.org)